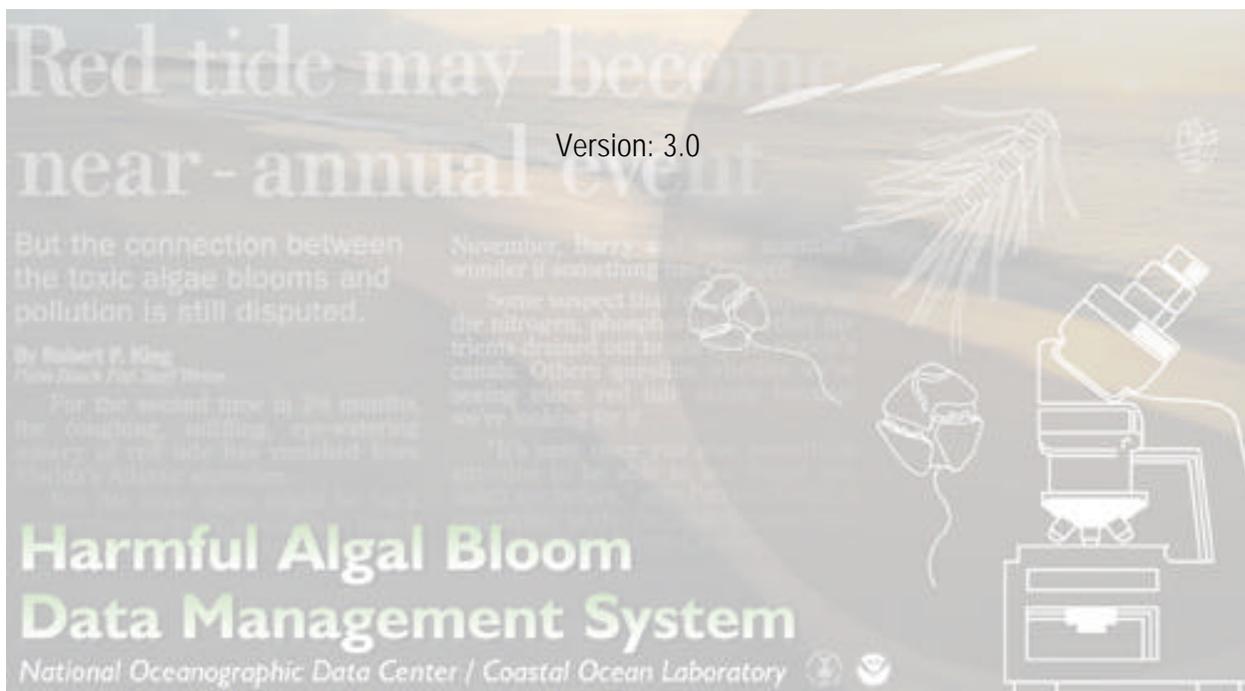


Harmful Algal Bloom Data Management System (HAB-DMS)

Database Design



HAB Team
National Oceanographic Data Center, NOAA
Silver Spring, Maryland
June 7, 2000

I. Introduction

The preliminary HAB database was designed to hold biological and chemical information at their most detailed level so that specific data collection and analysis methods can be stored using a data ingest program and queried and retrieved using a data access program. Oceanographic parameters which will be managed within the HAB database for HAB events include water quality measurements associated with toxic species abundance measurements, toxin, fish kill and phytoplankton distribution data. Time series and profile measurements collected separately from specific HAB monitoring programs, will be provided through the HAB-DMS, although they will physically be managed withing NODC's ocean profile and time series databases. Code tables are being developed wherever possible, so that data loading is controlled.

II. Hardware/Software Requirements

- A. Platform: SGI Origin 2000
- B. Operating System: Irix 6.5
- C. Software: Oracle 8.0.3.0.0, Data Direct Connect Open Database Connectivity (ODBC) driver, Microsoft Access, SQL*Net

III. Database terminology used within this document

A. Database terminology

1. Primary key: the column (field) or combination of columns that uniquely identifies a row.
2. Foreign key: a column or combination of columns whose values match the primary key. A foreign key doesn't have to be unique. They are often in a many-to-one relationship to a primary key.
3. Primary table: It contains the bulk of the data in the database and are generally related to one another by multiple key fields
4. Lookup table: It holds code definitions and/or lists of accepted values. It is usually one table in one to many relationships and data (row) is predefined and generated before primary tables.

B. Typographical conventions used to distinguish between primary keys, foreign keys and required fields:

1. **column*** - denotes a primary key column, there can be more than one column.
2. **column** - denotes a foreign key column

3. column⁺- denotes a required column

IV. Primary Tables

Table: Event

1. *Description:* Each row of this table contains information pertaining to where, when, and how a sample was taken, in the most general sense. Data submitted to the NODC may come in various time and spatial scales. One submission is considered an accession and gets a unique number for archival purposes. Within the database, the EVENT table is used to summarize some general information regarding the accession. Since this table is the highest in the hierarchy, a row of event information should be generated before any other tables except lookup tables. Each event is identified by an auto-generated unique number field, “event_no”.

2. *Relations to other tables:*

- a. Event : cruise_no - Cruise : cruise_no (many-to-one)
- b. Event : event_no - Environment : event_no (one-to-many)
- c. Event : event_no - Collection : event_no (one-to-many)
- d. Event : project_no - Project : project_no (many-to-one)
- e. Event : investigator_no - Investigator : investigator_no (many-to-one)
- f. Event : submitter - Investigator : investigator_no (many-to-one)
- g. Event : region - Region : region (many-to-one)
- h. Event : site_type - Site_Type : site_type (many-to-one)
- i. Event : z_unit - Unit : unit_no (many-to-one)

Name	Format	Description
event_no* ⁺	number	auto-generated unique number
<i>project_no</i>	number	foreign key - table: Project
<i>cruise_no</i>	number	foreign key - table: Cruise
<i>investigator_no</i>	number	foreign key - table: Investigator
source	number	foreign key - table: Investigator, the person who submitted data
<i>region</i>	char	foreign key - table: Region
<i>site_type</i>	char	foreign key - table: Site_Type

accession_no ⁺	char	accession number assigned by NODC
station	char	sampling station identifier
start_date	date/time	event start date, YYYYMMDD
end_date	date/time	event end date, YYYYMMDD
<i>remark</i>	memo	remark, other miscellaneous information entered during ingest

Table: Collection

1. *Description:* This table describes methods used at the time of collection as well as where and when the sample was taken. There are one or more collections per event. Each collection is uniquely identified by an auto-generated number, “collection_no” combined with “event_no”.

2. *Relation to other tables:*

- a. Collection : event_no - Event : event_no (many-to-one)
- b. Collection : event_no, collection_no - Biology : event_no, collection_no (one-to-many)
- c. Collection : event_no, collection_no - Chem_Anal : event_no, collection_no (one-to-many)
- d. Collection : z_unit - Unit : unit_no (many-to-one)
- e. Collection : gear_no - Gear : gear_no (many-to-one)
- f. Collection : ll_datum - LL_Datum : ll_datum (many-to-one)

Name	Format	Description
<i>event_no</i> * ⁺	number	foreign key - table: Event
collection_no* ⁺	number	auto-generated unique number per collection_no
<i>station_no</i>	number	foreign key - table: Station
start_date	date/time	date/time of collection
end_date	date/time	date/time of collection
tow_distance	number	distance of tow/haul in meters
latitude	char	latitude of collection location +/-DD.DDDDD (- is south)
longitude	char	longitude of collection location +/-DDD.DDDDD (- is west)

<i>ll_datum</i>	char	foreign key - table: LL_Datum
upper_depth	number	upper depth sample collection
lower_depth	number	lower depth sample collection
total_depth	number	total collection depth
<i>z_unit</i>	number	foreign key - table: Unit, depth unit
layer	char	foreign key - table: Layer
<i>gear_no</i>	number	foreign key - table: Gear

Table: Environment

1. *Description:* This table describes weather-related information, such as wind speed and direction, precipitation, water color, etc. There will be one record for every weather-related parameter measured during collection. Each measurement is uniquely identified by a “parameter_no” combined with “collection_no” and “event_no”.

2. *Relation to other tables:*

- a. Environment : event_no - Event : event_no (many-to-one)
- b. Environment : event_no, collection_no - Environment : event_no, collection_no (one-to-many)
- c. Environment : unit_no - Unit : unit_no (many- to-one)
- d. Environment : remark_no - Remark : remark_no (many-to-one)

Name	Format	Description
<i>event_no</i> ^{*+}	number	foreign key - table: Event
<i>collection_no</i> ^{*+}	number	foreign key - table: Collection
<i>parameter_no</i> ^{*+}	number	foreign key - table: Parameter
<i>unit_no</i> ⁺	number	foreign key - table: Unit
value ⁺	number	parameter value

<i>remark_no</i> ⁺	number	foreign key - table: Remark
--------------------------------------	--------	-----------------------------

Table: Chem Anal

1. *Description:* This table contains information about the chemical analysis of a sample within the laboratory. There can be one or more rows per collection.

2. *Relation to other tables:*

- a. Chem_Anal : collection_no, collection_no - Collection : event_no, collection (many-to-one)
- b. Chem_Anal : paramter_no - Paramter : paramter_no (many-to-one)
- c. Chem_Anal : unit_no - Unit : unit_no (many-to-one)
- d. Chem_Anal : method_no - Method : method_no (many-to-one)
- e. Chem_Anal : remark_no - Remark : remark_no (many-to-one)
- f. Chem_Anal : institute_code - Institute : institute_code (many-to-one)
- g. Chem_Anal : analyst - Investigator : investigator_no (many-to-one)

Name	Format	Description
<i>event_no</i> ^{*+}	number	foreign key - table: Event
<i>collection_no</i> ^{*+}	number	foreign key - table: Collection
<i>sample_no</i> ^{*+}	char	auto-generated unique number per collection_no
<i>parameter_no</i> ^{*+}	number	foreign key - table: Parameter
<i>sample_type</i>	char	sample type
<i>orig_sample_code</i>	char	originator's sample code/number
<i>value</i> ⁺	number	parameter value
<i>unit_no</i> ⁺	number	foreign key - table: Unit
<i>qualifier</i>	char	parameter detection limit qualifier code
<i>lab</i>	number	foreign key - table: Institute
<i>investigator_no</i>	number	foreign key - table: Investigator
<i>method_no</i>	number	foreign key - table: Method

<i>remark_no</i>	number	foreign key - table: Remark
-------------------------	--------	-----------------------------

Table: Biology

1. *Description:* This table contains information regarding biological samples that are further analyzed within a laboratory. This is a child table of Collection table, where these two tables are related by event_no and collection_no fields (columns). Sample_no, parameter_no, and tsn_no fields combined makes a unique row per collection.

2. *Relation to other tables:*

- a. Biology : event_no, collection_no - Collection : event_no, collection_no (many-to-one)
- b. Biology : parameter_no - Parameter : parameter_no (many-to-one)
- c. Biology : tsn - Taxonomy : tsn (many-to-one)
- d. Biology : method_no - Method : method_no (many-to-one)
- e. Biology : investigator_no - Investigator : investigator_no (many-to-one)
- f. Biology : remark_no - Remark : remark_no (many-to-one)
- g. Biology : bio_group_no - Bio_group : bio_group_no (many-to-one)

Name	Format	Description
<i>event_no</i> * ⁺	number	foreign key - table: Event
<i>collection_no</i> * ⁺	number	foreign key - table: Collection
sample_no* ⁺	char	originator's sample number
<i>parameter_no</i> * ⁺	number	foreign key - table: Parameter
<i>tsn</i> * ⁺	number	foreign key - table: Taxonomy
sample_type	char	sample type
orig_sample_code	char	originator's sample code/number
<i>investigator_no</i>	number	foreign key - table: Investigator
<i>method_no</i>	number	foreign key - table: Method
<i>bio_group_no</i>	number	foreign key - table: Bio_Group
<i>remark_no</i>	number	foreign key - table: Remark

Table: Bio_Desc

1. *Description:* This table contains detailed information per biological object, regarding taxonomy, life stage, etc. This is where the actual parameter value is stored (cell count, weight, etc). It is the last table in hierarchy, and its parent table is Biology.

2. *Relation to other tables:*

A. Bio_Desc : event_no, collection, sample_no, parameter_no, tsn_no - Biology : event_no, collection, sample_no, parameter_no, tsn_no (many-to-one)

F. Bio_Desc: life_stage_code - Life_Stage : code (many-to-one)

G. Bio_Desc: sex_code - Sex : code (many-to-one)

I. Bio_Desc: unit_no - Unit : unit_no (many-to-one)

Name	Format	Description
<i>event_no</i> * ⁺	number	foreign key - table: Biology
<i>collection_no</i> * ⁺	number	foreign key - table: Biology
<i>sample_no</i> * ⁺	char	originator's sample number
<i>parameter_no</i> * ⁺	number	foreign key - table: Biology
<i>tsn_no</i> * ⁺	number	foreign key - table: Biology
<i>desc_no</i> * ⁺	number	biological description number, auto-generated, unique per row of Biology table
modifier	char	taxonomic modifier (eg., sp, spp)
<i>life_stage_code</i>	number	foreign key - table: Life_Stage
<i>sex_no</i>	number	foreign key - table: Sex
value	number	value
<i>unit_no</i>	number	foreign key - table: Unit
remark	memo	remark, other miscellaneous information entered during ingest

V. Lookup Tables

Table: Cruise

1. *Description:* It contains general information about the cruise regarding overall time period, researchers, platform, or projects. Refer to the NODC master codes for “platform” field.

2. *Relations to other tables:*

- a. Cruise : cruise_no - Event : cruise_no (one-to-many)
- b. Cruise : project_no - Project : project_no (many-to-one)
- c. Cruise : Institute_code - Institute : institute_code (many-to-one)

Name	Format	Description
cruise_no* ⁺	number	auto-generated unique number
cruise_id ⁺	char	cruise id generated/provided by source
start_date	date/time	cruise start date
end_date	date/time	cruse end date
chief_scientist	char	name of person in charge of cruise
platform	char	platform (i.e., vessel/aircraft/vehicle name)
<i>project_no</i>	number	Foreign key - table: Project
<i>institute_code</i>	number	Foreign key - table: Institute

Table: Taxonomy

1. *Description:* This table contains general information about taxonomic identification. This table may be linked to the Integrated Taxonomic Information System (ITIS) in the future.

2. *Relation to other tables:*

- a. Taxonomy : tsn - Biology : tsn (one-to-many)

Name	Format	Description
------	--------	-------------

tsn* ⁺	char	identical to ITIS serial numbers
nodc_code	char	old NODC code, if any
sci_name	char	scientific (Latin) name
tsn_level	char	taxonomic level (i.e., genus, species)

Table: Remark

1. *Description:* This table contains information about remarks encountered during measurement of the parameter value.

2. *Relations to other tables:*

- a. Remark : remark_no - Chem_Anal : remark_no (one-to-many)
- b. Remark : remark_no - Biology : remark_no (one-to-many)

Name	Format	Description
remark_no* ⁺	number	auto-generated unique number
description ⁺	char	remark description

Table: LL Datum

1. *Description:* This table contains latitude/longitude datum codes and descriptions. The code defines the datum under which the latitude and longitude measurements for a particular station were calculated.

2. *Relations to other tables:*

- a. LL_Datum : ll_datum - Collection : ll_datum (one-to-many)

Name	Format	Description
ll_datum* ⁺	char	geographic datum associated with latitude and longitude values
description ⁺	char	description of datum code values

Table: Site Type

1. *Description:* Contains information relating to the criteria for sampling site selection. (i.e., fixed location sampling site, fish kill site, randomly selected sampling site)

2. *Relation to other tables:*

a. Site_type : site_type - Event : site_type (one-to-many)

Name	Format	Description
site_type* ⁺	char	sample site type code
description ⁺	char	sampling site description

Table: Region

1. *Description:* This table contains commonly used geographical terms (i.e., Gulf of Mexico)

2. *Relation to other tables:*

a. Region : region - Event : region (one-to-many)

Name	Format	Description
region* ⁺	char	unique code for a region or area
description ⁺	char	description of region
s_lat	char	southernmost latitude
n_lat	char	northernmost latitude
w_lon	char	westernmost longitude
e_lon	char	easternmost longitude
<i>ll_datum</i>	char	foreign key - table: LL_Datum

Table: Parameter

1. *Description:* This table contains information about parameters.

2. Relation to other tables:

- a. Parameter : parameter_no - Biology : parameter_no (one-to-many)
- b. Parameter : parameter_no - Chem_Anal (one-to-many)
- c. Parameter : parameter_no - Environment (one-to-many)

Name	Format	Description
parameter_no* ⁺	number	auto-generated unique number
parameter_code ⁺	char	parameter code
description ⁺	char	parameter description

Table: Institute

1. *Description:* Contains general information about institutes such as labs, research centers, colleges, government agencies, etc.

2. Relations to other tables:

- a. Institute : institute_code - Investigator : current_institute (one-to-many)
- b. Institute : institute_code - Project : institute_code (one-to-many)
- c. Institute : institute_code - Cruise : institute_code (one-to-many)

Name	Format	Description
institute_no* ⁺	number	auto-generated unique number
name ⁺	char	institute name, usually acronym
description	char	description
director	char	person in charge
street1	char	street
street2	char	street (continued)
city	char	city
state	char	state

country	char	country
zip	char	zip code
phone	char	phone number
fax	char	fax

Table: Investigator

1. *Description:* Contains general information about people such as investigators, researchers, scientists, etc.

2. *Relations to other tables:*

- a. Investigator : investigator_no - Event : investigator_no (one-to-many)
- b. Investigator : investigator_no - Biology : investigator_no (one-to-many)
- c. Investigator : current_institute - Institute : institute_code (many-to-one)

Name	Format	Description
investigator_no* ⁺	number	auto-generated unique number
last_name ⁺	char	last name of investigator
first_name	char	first name of investigator
middle_name	char	middle name of investigator
street	char	street
city	char	city
state	char	state
zip	char	zip code
country	char	country
phone	char	phone number
fax	char	fax
email	char	email

<i>current_institute</i>	number	foreign key - table: Institute
--------------------------	--------	--------------------------------

Table: Project

1. *Description:* Contains general information about projects and programs.

2. *Relations to other tables:*

- a. Project : institute_code - Institute : institute_code (many-to-one)
- b. Project : pi - Investigator : investigator_no (many-to-one)
- c. Project : project_no - Event : project_no (one-to-many)
- d. Project : funding_agency - Institute : institute_code (many-to-one)

Name	Format	Description
project_no**+	number	auto-generated unique number
name ⁺	char	project name, usually acronym
description	char	project description
<i>institute_no</i>	number	Foreign key - table: Institute, the institute which is responsible for the project
<i>pi</i>	number	Foreign key - table: Investigator
grant_no	char	grant number
grant_start_date	date/time	grant start date
grant_end_date	date/time	grant end date
<i>funding_agency</i>	number	Foreign key - table: Institute, funding agency

Table: Bio_Group

1. *Description:* It is a lookup table for biological groups, such as “phytoplankton”, “algae”, “fish”, where taxonomic information is not available.

2. *Relation to other tables:*

- a. Bio_Group : bio_group_no - Biology : bio_group_no (one-to-many)

Name	Format	Description
bio_group_no*	number	auto-generated unique number
description ⁺	char	bio group description

Table: Life Stage

1. *Description:* It is a lookup table for taxonomic life stages.

2. *Relation to other tables:*

a. Lifestage : code - Bio_desc : life_stage_code (one-to-many)

Name	Format	Description
life_stage_no**	number	auto-generated unique number
description ⁺	char	life stage description

Table: Sex

1. *Description:* It is a lookup table for taxonomic sex.

2. *Relation to other tables:*

a. Sex : code - Bio_Desc : sex_code (one-to-many)

Name	Format	Description
sex_no**	number	auto-generated unique number
description ⁺	char	sex description

Table: Gear

1. *Description:* It is a lookup table for gears used for collection

2. *Relation to other tables:*

a. Gear : gear_no - Collection : gear_no (one-to-many)

Name	Format	Description
gear_no**+	number	auto-generated unique number
model ⁺	char	gear model/type
description	char	trophic level description

Table: Unit

1. *Description:* It is a lookup table of units for measurement.

2. *Relation to other tables:*

a. Unit : unit_no - Collection : unit_no (one-to-many)

Name	Format	Description
unit_no**+	number	auto-generated unique number
unit_code ⁺	char	unit
description	char	unit description

Table: Station

1. *Description:* This table contains exact location information for stations associated with routine monitoring programs.

2. *Relation to other tables:*

a. Region : region - Event : region (one-to-many)

Name	Format	Description
station_no**+	char	unique code for station
station_id ⁺	char	originator's collection code

latitude	char	latitude
longitude	char	longitude
<i>ll_datum</i>	char	foreign key - table: LL_Datum
program	number	foreign key - table Project, name of project stations are associated with
remark	memo	remark, other miscellaneous information entered during ingest

Table: Method

1. *Description:* It contains descriptions of the field and laboratory methods for parameter determination. (similar to “BIO_METHODS” table of CBP)

2. *Relation to other tables:*

a. Parameter : parameter_no - Method: parameter_no (one-to-many)

Name	Format	Description
method_no* ⁺	number	auto-generated unique number
method_code ⁺	char	analytical method code number
type ⁺	char	method type or other identifier
description ⁺	char	method description
reference	char	bibliographical reference for published methods
reporting_limit	number	value of the methods reporting limit
parameter_no	number	foreign key - table: Parameter, parameter associated with method

June 7, 2000 (Version 2.0.2)

HAB Specific Definition Tables

Primary Tables

NODC Definition Tables

Primary Key (PK)

Foreign Key (FK)

Project
project_no
name
description
institute_no
pi
grant_no
grant_start_date
grant_end_date
funding_agency

Institute
institute_no
name
description
director
street1
street2
city
state
country
zip
phone
fax

Cruise
cruise_no
cruise_id
start_date
end_date
chief_scientist
platform
project_no
institute_no

Investigator
investigator_no
last_name
first_name
middle_name
street1
street2
city
state
zip
country
phone
fax
e-mail
current_institute

Bio_group
bio_group_no
description

Gear
gear_no
model
description

LL_datum
ll_datum
description

Site_type
site_type
description

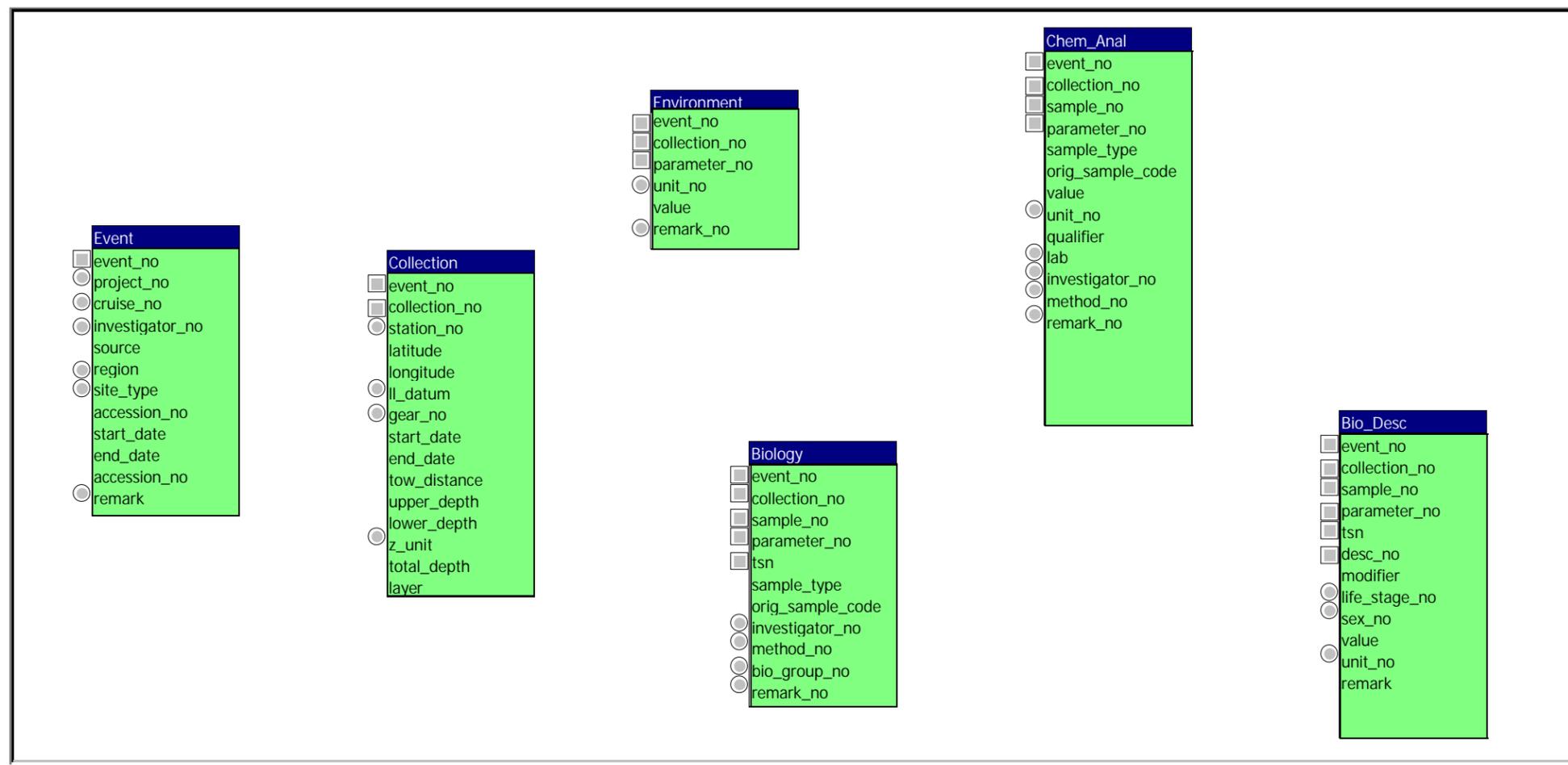
Unit
unit_no
unit_code
description

Remark
remark_no
description

Taxonomy
tsn
nodc_code
sci_name
tsn_level

Parameter
parameter_no
parameter_code
description

Region
region
description
s_lat
n_lat
w_lon
e_lon
ll_datum



Station
station_no
station_id
latitude
longitude
ll_datum
program
remark

Method
method_no
method_code
type
description
reference
reporting_limit
parameter_no

Life_Stage
life_stage_no
description

Sex
sex_no
description